DOCUMENT RESUME

ED 336 954 FL 019 477

AUTHOR Abbey, Susan
TITLE A Case for On-Going Evaluation in English Language Teaching Projects. CLCS Occasional Paper No. 29.
INSTITUTION Trinity Coll., Dublin (Ireland). Centre for Language and Communication Studies.
REPORT NO ISSN-0332-3889
PUB DATE 91
NOTE 47p.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *English (Second Language); Evaluation Methods; Foreign Countries; Instructional Innovation; *Program Descriptions; Program Design; *Program Evaluation; Program Implementation; *Second Language Instruction
IDENTIFIERS Malaysia

ABSTRACT

The first section of this paper considers what is understood by evaluation in the field of project design for English language teaching (ELT). Published accounts of ELT projects are examined for their policy on evaluation and a list of dominant themes in project evaluation is drawn up. The second section reviews theoretical issues involved in project evaluation and seeks to summarize what might be included in a system of evaluation conceived as being central to the design and implementation of a project. The third section discusses the idea that project designers need to take account of the impact of the innovations they introduce. Consideration is given to how this aspect of project design might influence the forms of evaluation employed by project designers. The fourth section looks at a particular ELT project from a teacher's point of view, considering the forms of evaluation employed in the project and the extent to which they match the features outlined at the end of the second section. The fifth section offers suggestions for planning the evaluation, and particularly the on-going evaluation, of projects in ELT design. (JL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED336954

TRINITY COLLEGE DUBLIN

Centre for Language and Communication Studies

A case for on-going evaluation in English language teaching projects

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Susan Abbey

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Abbey Susan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CLCS Occasional Paper No. 29

ISSN 0332 3889

Spring 1991

FL019477

BEST COPY AVAILABLE

CLCS Occasional Papers

General Editor: D. G. Little

CLCS Occasional Papers report on research carried out within or in association with the Centre for Language and Communication Studies, and on research carried out elsewhere which is of special interest to the Centre's own concerns and activities. In some instances they present the texts of public lectures promoted by the Centre.

The following titles have been published and are available from The Secretary, Centre for Language and Communication Studies, Trinity College, Dublin 2.

Autumn 1981

1. D. M. Singleton, *Language transfer: a review of some recent research* (31pp.)
2. Jeffrey L. Kallen, *Linguistics and oral tradition: the structural study of the riddle* (33pp.)
3. D. M. Singleton, *Age as a factor in second language acquisition* (70pp.) - OUT OF PRINT

Summer 1982

4. Francis Nolan, *Voice quality and speech synthesis* (30pp.)
5. Roger Bennett, *Language elicitation procedures* (23pp.)

Winter 1982-3

6. S. M. Devitt, D. G. Little, S. P. O Conchúir & D. M. Singleton, *Learning Irish with Anois is Arís* (128pp.) - OUT OF PRINT

Autumn 1983

7. William T. Littlewood, *A communicative approach to language-teaching methodology* (19pp.)
8. Rose Maclaran, *On the interaction of semantics and pragmatics* (18pp.)
9. E. M. Harding, *Compensation strategies* (54pp.)

Autumn 1984

10. Jeffrey L. Kallen, *Generative phonology in the clinic* (38pp.)
11. Gearóid O Ciaráin, *The affective dimension in second/foreign language learning: an interactional perspective* (42pp.)

Spring 1985

12. D. M. Singleton & D. G. Little, *Foreign languages at second level: defining a syllabus, with particular reference to the needs of the senior cycle* (19pp.)
13. John Harris, *The polylectal grammar stops here* (12pp.)

(Continued on inside back cover)

CLCS Occasional Paper No.29
Spring 1991

A case for on-going evaluation in English language teaching projects*

by

Susan Abbey

0 INTRODUCTION

Evaluation is a much debated term. Different people use it to mean different things, and different aspects of it are considered important by different people. The first section of this paper considers what is understood by evaluation in the field of project design for English language teaching (ELT). Published accounts of ELT projects are examined for their policy on evaluation and a list of dominant themes in project evaluation is drawn up. In the light of these themes, the second section goes on to review theoretical issues involved in project evaluation and seeks to summarize what might be included in a system of evaluation conceived as being central to the design and implementation of a project. It considers in particular the advantages of on-going evaluation. The third section of the paper discusses the idea that project designers need to take account of the impact of the innovations they introduce. Consideration is given to how this aspect of project design might influence the forms of evaluation employed by project designers. The fourth section looks at a particular ELT project from a teacher's point of view, considering the forms of evaluation employed in the project and the extent to which they match the features outlined at the end of the second section. Finally, drawing on the conclusions of the previous four sections, the fifth section offers some suggestions for planning the evaluation, and particularly the on-going evaluation, of projects in ELT design.

---

* An earlier version of this paper was submitted in fulfilment of the requirements of the M.Phil. in Applied Linguistics, Trinity College, Dublin, September 1989.

1 EVALUATION IN PROJECT DESIGN: A REVIEW OF PROJECT DESCRIPTIONS

This section examines the published descriptions of eight projects in order to obtain a general idea of what aspects of evaluation are raised by different writers working on different types of projects. The projects were carried out in a number of different countries and ranged from primary to postgraduate level, from institutional to national scale, from the writing of a course for a particular group of learners to course book writing, teacher training and research. The published descriptions all make some reference to project evaluation and attempt to incorporate it in a project plan in some form or other. We are particularly interested to discover:

- what importance is currently accorded to project evaluation;
- what aspects of project evaluation are considered important;
- what form the evaluation takes;
- how it is implemented.

1.1 Study Skills in English (Candlin, Kirkwood and Moore 1977)

This article reports on a three-week intensive course in English for students coming to Britain in order to take postgraduate courses. The aim was to design pre-sessional courses for the areas of Engineering, Economics, Urban Planning, and Foreign Service. The article describes the issues that arose in the planning and implementation of the course. Evaluation was built into the model of course design at the final stages in a planning procedure which consisted of six stages: objectives; syllabus content; course programme; teaching; evaluation of students; evaluation of course.

In the model, evaluation of students feeds back into the teaching component, while evaluation of the course feeds back into the objectives component. However, we are given no real indication as to how evaluation is to take place or who should be responsible for it, nor is there any description of a systematic evaluation procedure.

1.2 Education, ideology and materials design: a Tanzanian experience (Brumfit 1980)

This report describes a project in syllabus design, and ultimately materials design, in Tanzania. Brumfit emphasizes that a main feature of the project was the exchange of ideas between the curriculum development unit, the inspectorate, teachers in schools, and teacher trainers, with a view to obtaining regular feedback about the course and its progress in order to institute change. The idea was that this feedback and the subsequent change should be "principled rather than random" (Brumfit 1980, p.166).

The machinery provided to support this evaluation procedure was a

regular programme of in-service courses for teachers, the appointment of "co-ordinators" (teachers who taught and acted as advisers) in the various zones of the country, local independent English language teachers' associations, and a national organisation of English teachers which produced a bulletin. These bodies allowed for the discussion, criticism, and dissemination of ideas concerning the project and its evaluation.

A syllabus together with an outline of the principles behind it was circulated to all participants and acted as a working document. Teachers could follow it, or disagree with it and make their own changes, or take its ideas and develop them further. In sum, it provided a discussion document for all concerned. The intention of the project designers was that the document would be revised and improved as a result of the feedback coming through the channels specified above. Thus the document was seen as providing a framework for change and development rather than a fixed scheme.

1.3 Planning a project: the KELT project, Sierra Leone (Hayes 1983)

This project aimed to support and co-ordinate the production of a syllabus for English at primary level, materials to match, an entrance exam to reflect the syllabus, a teacher training syllabus based on the project, and an in-service scheme. The planning was based on three principles, the first of which was flexibility: "We tried not to strangle ourselves with our own project specifications, in the knowledge of what can so easily happen to the best-laid plans. Planning is an on-going process in the life of any project, in that it must respond to the inevitable changes in the situation. Our priority is a strong framework which can cope with any necessary changes of activity as the years go on, which can respond to any new need, which can survive when the expatriates move out" (Hayes 1983, p.25). We conclude from this that changes to the original project plan are desirable and that the framework should be designed to cope with change. In other words, evaluation should take place throughout the project and feed back into its structure. But how exactly this should happen we do not learn from the report.

1.4 The English Language Centre, Jeddah (Roe 1980)

This report describes the work of the English Language Centre, which provides English language courses for students of the colleges of medicine and engineering at the King Abdul Aziz University in Jeddah. Roe suggests that various methods of project evaluation can be used:

- comparison of the success of one programme with another in terms of students' results (expressed in bands);
- comparison between classes within courses;

- comparison between the progress of a group in one semester and its progress in previous semesters.

Clearly Roe is concerned chiefly with evaluation of student performance, and seems to judge the success of a course on this alone.

1.5 The Francophone Primary Project, Cameroon (Wilson and Harrison 1983)

This project aimed to introduce English into the last three classes in all primary schools in Cameroon. It involved the training of approximately 6,000 teachers through a one-year pre-service course and in-service courses. A major part of the project was the production of three textbooks.

The original project started in 1975, and according to Wilson and Harrison was not a success. They blame its failure largely on:

- poor initial project design;
- lack of consultation during initial planning;
- lack of observation of classroom procedures in the early stages of the project;
- lack of rigorous trialling of materials;
- the fact that the writers had no experience of teaching the materials;
- no close examination of teacher supply figures.

On the basis of this experience, Wilson and Harrison propose a model consisting of a series of steps necessary for successful project planning. Evaluation plays an important part in this at two stages:

(i) Pre-trialling of materials, where the writers themselves try out sample materials. The aim is to assess the suitability of particular activities; to calculate what might be covered in one lesson; to judge if student teachers could handle materials; and to weed out political and cultural errors.

(ii) Formal trialling of materials. Wilson and Harrison stress the importance of including all categories of teachers—trained and untrained, experienced and inexperienced, good and mediocre, those with and those without support.

Wilson and Harrison suggest that it is necessary to have a checklist of information required on the appropriacy and effectiveness of the materials. The methods they suggest for data collection are questionnaires after every five lessons, discussions with the teachers, and observation of teachers using the materials. A final aspect of evaluation that they mention is a post-course assessment of the materials. They feel that this should involve evaluation of pupil performance through criterion-referenced tests.

1.6 The English Language Textbook Project - Somalia (Clarke, Hawkes, Pritchard, and Smith 1983)

This project concerned the design of a textbook for secondary level English teaching and the writing of a scientific English course, plus a reader, for students undertaking vocational and technical training through the medium of English. The project also involved the setting up of pre-service and in-service training courses for teachers who were to use the textbooks.

The first attempt at evaluation was made during the writing of the materials. It was carried out by trainee teachers, who as part of their degree had to research and write papers on the textbook project. They looked at

- potential problems for Somali teachers and students with the new materials;
- the rationale behind the textbook;
- the economics of textbook production.

Clarke et al. make the point that systematic and on-going evaluation was desirable since this project represented a new direction in English language teaching in Somalia. Furthermore the projected lifetime of the textbook was 10-15 years. "Evaluation, then, is considered by the team to be a core activity in all stages of the project. Evaluated data provides or has provided the basis for judgements to be made on:

(a) the nature of product;
(b) the product quality and quantity;
(c) the operational success of the product;
(d) training requirements for serving teachers and teacher-trainees;
(e) the impact of the programme on English language teaching and learning in Somalia in the light of the project objectives" (Clarke et al. 1983, p.65)

The instruments of evaluation used in trialling were as follows:

- Reports of classroom observation on the ability of teachers to use the materials successfully. This should indicate areas in which pre-service training is necessary and assess student response to materials.
- Users' reports filled in by each teacher during trialling to provide data on pace and specific problem areas encountered. This should provide a lesson-by-lesson evaluation which indicates teachers' attitudes to new materials.
- A questionnaire filled in by teachers on completion of the trialling package to provide data on student motivation, teacher attitudes to teacher notes and materials, and specific problem areas.
- A questionnaire filled in by students to gauge their reaction to the new courses.
- Regular informal interviews with teachers based on their question-

naires and reports to elaborate on the data.
- Regular samples of students' work to evaluate their success in following the course.
- Tests based on the performance objectives of the textbook to assess the pedagogic success of the materials.

The resulting data was used:
- to improve the production of subsequent books;
- to improve the provision of pre/in-service teacher training;
- to improve support for the teacher through better teacher notes and guides;
- to assess the overall success of the project so as to inform projects of a similar nature elsewhere.

The constraints on this evaluation were as follows:
- There was a pressing need for materials, and this meant that the first book was not trialled prior to its publication. Therefore, information obtained from subsequent trialling was not used to change the materials but was fed into the teacher-training programme to help teachers cope with the flaws.
- Geographic distance and poor communications meant that it was difficult to monitor the implementation of the programme.
- Factors such as lack of adequate facilities (e.g., equipment and rooms) and inadequate training of teachers in the past had to be taken into account.

1.7 The Ain Shams University Project, Egypt (Bowers 1983)

The primary aim of this project was to provide undergraduate curriculum revision in TEFL, postgraduate TEFL training, test development, conference and seminar organization, and research activities. Bowers looks at how the mismatch between project planning and project performance can be minimized. The main thrust of his argument is that any change in a system will initiate changes in all other sectors of that system. This is what he refers to as the "spider's web" effect (Bowers 1983, p.100): when you touch one part, the whole vibrates. Therefore an allowance for change needs to be built into a project at the planning stage, and every project should be aware of its effect on other sectors.

Bowers (1983, p.115) suggests that the following operational procedures should be included in the planning of a project in order to cater for the spider's web effect:

(i) A problem or need is recognized.
(ii) The real problem or need is identified.
(iii) Possible solutions are considered.
(iv) The implications of each are followed through.

(v) A course of action is decided upon and this becomes the operational design.
(vi) Personnel and resources are negotiated.
(vii) All interested parties are involved.
(viii) Implementation commences.
(ix) Formative evaluation and regular reporting takes place.
(x) At an appropriate point a summative evaluation is carried out and the cycle is repeated/adjusted/terminated.
(xi) Experience is disseminated.

Bowers stresses the need for communication between the project team, the administrative and financing authorities, teachers, and trainers. In addition, the project team needs to communicate within itself, to acknowledge and incorporate different views. Therefore there must be "clarification of the criteria and processes for on-going evaluation" (Bowers 1983, p.116), and this needs to be done at the outset, and to be specific. At the same time it should be flexible to allow for adjustments resulting from policy changes or practical constraints. Bowers says that there is little to be gained in saving evaluation for the end of the project. Evaluation at this stage might show weaknesses, but it would not be possible to correct them; it might show strengths, but it would be too late to develop them.

On the question of who should be responsible for evaluation, Bowers (1983, p.117) points out the advantage of the implementer, e.g. teacher, playing the role of evaluator. He stresses the useful insights that an "insider", such as a teacher, would have, which the outsider, e.g. foreign expert, might not have. He further stresses the need to exploit the contribution that different parties can make to the curriculum process. Finally, he suggests (1983, p.117) that it is normal and desirable that project expectations and project results should differ, since curriculum development is a dynamic process, created by the interaction between innovation and the context.

1.8 The University of Nairobi Project (Brown and Hirst 1983)

This project was set up in the Language and Study Skills Unit of the University of Nairobi to provide an improved study skills course for students at the university. The main types of evaluation undertaken by the unit are:

(i) pre- and post-testing of courses written by the members of the unit in terms of the objectives of each course;
(ii) evaluation of the work of the unit in terms of its effect on student performance in their own subject areas.

According to Brown and Hirst, evaluation of a course requires a pre-test of two groups of students, those taking the course and a control group of

students not taking it. These groups should be matched in terms of linguistic skill, the subject courses they are taking, and initial subject knowledge. The post-test should be a subject test, devised by the subject lecturer. Brown and Hirst point out some constraints on this approach. For example, there are practical problems in getting a suitable control group and in co-ordinating the post-test with subject teachers. It is also difficult to determine whether improvement is due to the study skills course or simply to attendance at the university (Brown and Hirst 1983, p.143).

Brown and Hirst suggest that it is necessary to test not only whether students have learnt anything in study skills classes, but also if what they have learnt is useful and if they can transfer it successfully to their subject areas. They say that there are no objective criteria for evaluating this. The methods they suggest are (Brown and Hirst 1983, p.144):

(i) judgement by members of the study skills unit on the effectiveness of the course;
(ii) the opinion of students on the course elicited through a post-course questionnaire;
(iii) the opinions of subject staff elicited through
- consultation on written work,
- observation of study skills classes,
- informal assessment of the effects of the course on students' work in subject areas.

Brown and Hirst (1983, p.146) feel that evaluation is essential. However, they are unhappy about the unclear and unrigorous methods that are suggested here. They feel strongly that more research needs to be done in this area to produce more satisfactory methods.

1.9 Conclusion

The eight projects presented above were chosen because of the variety of settings, aims, and scope they display. They all deal with evaluation in some form or other, and they are all relatively recent and therefore demonstrate up-to-date practice in this area. Collectively they confirm that evaluation is a very important aspect of project design and one that all project planners have to consider. We suggest that the following nine themes, all of which recur in these descriptions, are central questions in this debate, are to some degree unresolved, and are therefore worth further investigation:

(i) Why include evaluation in project design?
(ii) What should be evaluated?
(iii) What criteria should be used for evaluation?
(iv) Who should be involved in evaluation?
(v) What form should evaluation take?

(vi) Should evaluation be summative, formative, or both?
(vii) At what stages in the programme should evaluation take place?
(viii) What channels should be used to feed information back into the programme?
(ix) What kind of constraints can we expect to encounter in the evaluation of a programme?

2 EVALUATION: SOME THEORETICAL CONSIDERATIONS

We continue our survey of the literature by turning to three writers, Potts (1985), Murphy (1985) and Rea (1983 and 1987), who deal more specifically with theoretical aspects of project evaluation. We shall examine their views to discover their positions vis-à-vis the nine themes listed at the end of the first section; to discover what kind of arguments and polemics surrounding these issues need to be taken into consideration; and to establish what kind of theoretical stance we might adopt on these points.

2.1 Why include evaluation in project design?

Potts (1985) divides reasons for communicative curriculum evaluation into two types: internal, i.e. those generated by the nature of the communicative curriculum itself; and external, i.e. those designed to provide evidence to outside interests about what is happening on the course. Potts suggests that the first internal reason for a formal evaluation component in a communicative curriculum is "to sort out priorities, weight them and so guide future activity" (1985, pp.24f.). A second reason he gives is to allow the growth of a sense of achievement in the learner. Examples of external reasons for evaluation may be course validity, assessment of the progress of individual students, placement criteria for future courses, and assessment of the usefulness of materials.

The development of the communicative curriculum regards the question of how people learn to use a language as part of the teaching/learning process. Both teacher and learner are thus engaged in on-going formative research (Potts 1985, p.21). Central to this process are the needs and wants of the learner. Potts makes the point that unlike structural and functional curricula, where the the content of a curriculum is based on an inventory of items specified in advance, the communicative curriculum should allow for systematic change to its content. The evaluation procedure of a communicative curriculum, therefore, will measure shifts in learners' needs and wants and translate these into learning activity.

Murphy (1985, p.4) comments: "Evaluation should be an integral part of the working of the curriculum to ensure that what is done is worthwhile, necessary and sufficient." Like Potts, Murphy goes on to say that the

curriculum should not be regarded as static, and that evaluation may be used as a basis for further curricular change. Thus evaluation "contributes to adjusting the curriculum as needed, and to ensuring its smooth working" (Murphy 1985, p.4).

Evaluation is one way to produce information on assessment. This involves not only tests but assessment of all other elements of the curriculum. Furthermore, it is a means of producing information for public accountability. This means being able to report accurately on teaching, which in turn means reporting on good and bad results. A further reason for evaluation is to find out more about what actually happens in teaching and learning, and thus improve curriculum development. It is therefore a means of showing how well or badly a new theory of syllabus design works in practice. So what it requires is an empirical approach to new ideas on teaching and learning (Murphy 1985, p.9).

At the moment, the problem is that there is no formal means for teachers to find out about the effectiveness of their teaching in an empirical way, apart from testing. There are no indications of how teachers should interpret the results they are getting, and no proposals for alternative forms of evaluation that teachers could use for this purpose. Murphy criticizes syllabus designers who show a "give it a whirl and see how it goes" attitude to evaluation (1985, p.8). He takes the line that for too long we have been happy to rely on reasoned theory for the justification of our syllabuses, and for too long have ignored the implications of data from the classroom situation. His proposal, therefore, is for a comprehensive, balanced evaluation, which involves teachers at classroom level.

Rea (1983, p.89) gives the purposes of evaluation as:

(i) information-gathering and dissemination;
(ii) quality control;
(iii) accountability;
(iv) progress and achievement.

To allow people involved in the project design process to make effective decisions, it is necessary to obtain a certain amount of information at the various stages of the project. So, for example, materials writers will want feedback on the success or failure of their work, institutions will want information on student achievement, and sponsors will want information on the safeguarding of their investment. A further reason Rea gives for evaluation is to monitor what she refers to as the "localisation process" (Rea 1983, p.93), in projects where there is expatriate involvement.

According to Rea (1987, p.150) evaluation in project design has essentially two purposes. The first purpose is similar to Potts's internal reasons, and is intended to provide guidance for subsequent teaching and learning by using information derived at different intervals within the teaching and

learning process. The second purpose is to provide information on the adequacy of the curriculum in use to outside interests, such as clients who have commissioned the programme. This closely parallels Potts's external reasons for evaluation. Rea (1987, p.149ff.) suggests that a further important reason for the inclusion of evaluation in the design of a communicative curriculum is the need to validate new, theoretically appealing approaches.

There seems to be general agreement in the literature that evaluation is important for:

- validation of the curriculum design and of the theoretical concepts behind it;
- internal curriculum development;
- external face validity.

There is general dissatisfaction with the methods of evaluation currently in use, and a belief that there has been too much stress on evaluation for the purpose of assessment of student achievement. In general, therefore, the whole purpose of evaluation needs to be broadened, and this will have implications for all aspects of assessment.

2.2 What should be evaluated?

All three writers under review take the line that what is often assumed to be the main concern of evaluation, viz. learner performance, is by no means the only thing that should be evaluated. Potts (1985, p.22) suggests that there must be some evaluation of whether what is being learned is appropriate to the learners' needs; whether skills are being transferred; and if they are not, what range of strategies are necessary for successful transfer. Commenting on the idea of measuring students' progress in terms of a fixed target repertoire, Potts says (1985, p.20) that this "inventory" approach to evaluation is not really possible in a communicative curriculum because of the flexible and variable view of language on which it is based. Mastery of target language structures and vocabulary is not a sufficient measure of course success or student performance. Potts (1985, pp.20f.) makes the point that the communicative curriculum regards the question of discovering how people learn a language as part of the language learning process, which means that this too becomes the subject of evaluation to the extent that we wish to evaluate the process of language learning.

The demands of communicative methodology imply that what is learned and taught should be "real" language. Thus for projects in communicative language teaching a further subject for evaluation is whether the project includes such language (Potts 1985, p.21). Another demand made by communicative methodology is that the real needs of the learner should be met. This requires an emphasis on language in use, rather than on grammar or communicative functions. We would therefore want to evaluate the

learners' use of language in the classroom, rather than mastery of forms. We will also need to consider the extent to which classroom learning is related to real communicative needs, as well as the extent to which learners are enabled to bring existing communicative competence to the classroom. The point was made in the last section that learners need to be aware of their own progress. This has implications for what is evaluated and would require comparison of learners' previous product with their current product.

Murphy (1985, p.4) agrees with Potts that it is not just the performance of students that we should take as the subject of evaluation, but the various factors involved in the teaching and learning process. He details these as:

- the curriculum itself and the context in which it is set;
- materials;
- teaching and learning;
- attitudes;

Under teaching and learning we might examine individual learner performance through tests, but would also needs look at the different techniques and strategies of learning. Not all of these factors will be of equal importance for each type of evaluation, but whatever weighting each factor receives, it should be present. Furthermore there is a need to report on positive data as well as negative.

Rea (1983, p.90) also stresses the need to evaluate both the process and the product of learning. She makes the point that what is evaluated depends on what the purpose of the evaluation is and for whom and when it is conducted. For example, the data might be descriptive and serve an explanatory function. It might also be used to serve as a feedback to learners and teachers and act as a guide to changes in the future.

Rea (1987, p.150) talks about using evaluation to validate a curriculum. Here she firstly suggests that materials and tasks should be examined to establish whether they meet the original curriculum specifications and the needs of the target audience. Secondly she suggests looking at how the principles of the curriculum are realized in the actual learning units. She argues that a main function of evaluation is to assess the general suitability of the curriculum for the needs and demands for which it was designed. Rea notes that the emphasis in the evaluation of a communicative curriculum has to shift from content materials to learning activities and to the opportunities presented to learners for the development of their communicative abilities. Thus the things that will be evaluated will refer more to the process of learning than to the product.

On the whole our three authors feel that in the past there has been an exclusive concern with the course product. While this may still be of some relevance, the most important focus in a communicative curriculum must be on the process of learning and teaching.

2.3 What criteria should be used for evaluation?

According to Potts (1985), there are different types of criteria that can be used for evaluation. As we have already seen, he makes a broad distinction in evaluation between process and product. Product, Potts suggests, might be further divided into end-product and on-going product. Any of these aspects of a project can be examined in terms of four different criteria:

- initial input;
- target performance;
- the learner's previous performance;
- other learners' performance (Potts 1985, p.34).

Depending on the context of the project, more or less weighting might be given to particular types of criteria. For example, for an English for Specific Purposes (ESP) course it is quite likely that the end-product, and consequently target performance criteria, will be important. But the weighting of the different criteria will also depend on the perspective of the evaluator. For example, on an English for Academic Purposes (EAP) course, a subject teacher would not have the same perspective on evaluation as an EAP teacher and would consequently give a different weighting to different criteria. In any case, it is important that the criteria are clear to everybody involved, including the learner.

Murphy (1985, p.4) argues that the criteria will depend on whether the evaluation is concerned with the product of a course or with its processes. If it is with the former, then effectiveness will be the chief criterion, whereas if it is with the latter, the chief criterion will be efficiency. The effectiveness of product refers to the matching of the actual results of tests with the target results. Murphy points out that this criterion for evaluation is of limited usefulness for the improvement of a course since it does not tell us anything about the reasons for the discrepancy between expected and actual results. Nor does it give us any indication of ways of pin-pointing what it is that needs adjusting in order to minimize that discrepancy. Murphy (1985, p.4) goes so far as to suggest that this lack of information produces professional secrecy and entrenchment among teachers. Where efficiency of process is the chief criterion for evaluation, the focus is far less fixed. Murphy (1985, p.4) argues that this criterion "is more likely to involve all the groups concerned, to encourage them to cooperate and thus to produce the information on which to base development".

Rea (1983, p.91) stresses that no matter what, for whom, or when we wish to evaluate, it is the context that is the most important factor in determining the criteria for evaluation. Consequently, flexibility and responsiveness must be important factors in establishing criteria. As much as possible should be predicted about the evaluation procedures of any project, and planned into the design document. However, it is not possible to predict all

variables.

Examples of the types of criteria that Rea (1987, p.164) gives are:

- the diagnostic nature of tasks;
- the coherence of tasks within the social structure of the classroom;
- the complexity of tasks;
- how the success of tasks is determined;
- the transfer of skills from the classroom to the outside world;
- the extent to which activities in the classroom reflect the real needs of the students.

Criteria for evaluation are generally thought to be closely related to context. Consequently it is difficult to generalize about criteria in advance. However, it is considered an advantage to set out in advance as far as possible, not only what is going to be evaluated, but also some criteria for that evaluation. And it seems very important that all involved should be aware of what these criteria are.

2.4 Who should be involved in the evaluation?

In Murphy's view (1985, pp.2f.) one aim of evaluation is to develop the curriculum, so that anyone involved in its operation can make a valid contribution to its evaluation. This includes curriculum planners, teachers, learners, and people outside the institution, such as sponsors and society at large. Murphy is particularly concerned to emphasize the potential role of teachers, who in their function as curriculum managers are concerned with the same issues as the curriculum designers. He points out that their judgements may be as influential in the process of education as the formal findings of the experts. Evaluation, as it exists at the moment, does not exploit this potential. This is because currently available models of evaluation do not provide a suitable framework for the inclusion of teachers. Murphy (1985, pp.15f.) also seems to suggest that teachers do not have the necessary training to assume responsibility for evaluation.

Potts (1985, p.33), on the other hand, emphasizes the fact that on-going evaluation requires teachers and students together to be engaged in active research. He suggests (Potts 1985, p.23) that there is a tension between the need to teach languages authentically and the need to monitor and evaluate. For this reason the burden of evaluation is too great for the teacher to bear alone. If evaluation is to work, the learners too must be aware of and involved in the evaluation process. Potts (1985, p.38) further stresses the importance of suitable materials in this process. He points out the potential for confusion over the way in which teachers evaluate learners' performance if learners are not aware of the criteria teachers are using and if they remain outside the process.

Rea (1983, p.90) also stresses the range of people who may be involved

in evaluation to the extent that they may obtain information from it on which to make decisions—e.g. tutors, materials writers, host institutions, backing organisations, or governments. On the other hand, she argues (1983, p.90) that evaluation can be damaged if the detailed planning stages are handled by inexperienced people. The main danger is that they will impose too rigid a structure and not allow enough flexibility. However, in terms of who provides data for an evaluation, Rea is of the view that teachers and curriculum developers both have roles to play, as well as students and sources external to the language programme, for example, in the EAP setting, subject specialists (Rea 1987, p.159).

The various points of view we have been reviewing imply that traditionally evaluation is the domain of the expert and that therefore there are not many people who know much about it. It seems to be desirable to widen the range of people involved. It also seems important that the frameworks for evaluation should be more easily accessible to the non-expert. In sum, there is a call for the "demystifying" of evaluation.

2.5 What form should evaluation take?

Murphy (1985, p.15) suggests that there are two basic forms that evaluation can take:

(i) evaluation that produces qualitative results;
(ii) evaluation that produces quantitative results.

Whatever method is used, the starting point must be to know what kind of information is needed and what are the most appropriate ways of getting that information. In order to be appropriate, a method must be reliable, i.e. it must produce the same result if repeated in similar circumstances; it must be valid in the sense that it answers the questions we set out to answer; and it must be accessible in the sense that it can be described and replicated.

These requirements could be difficult to meet in the classroom and this would naturally have implications for Murphy's aim of involving teachers in evaluation. Accordingly, a clearly defined form for the evaluation is essential, in combination with procedures for familiarizing teachers with it.

We have already mentioned a problem associated with teacher evaluation of learner performance, namely possible confusion in the way learners interpret teachers' evaluation criteria. Potts (1985, p.33) is concerned that this problem should be solved and suggests that one way of doing this would be to put evaluation in the hands of the learners. This would obviously have consequences for the form that evaluation might take. For instance, it means that evaluation criteria must be learner-centred. To translate this objective into practicable forms and methods, usable by learners and teachers, could pose its own problems. Potts (1985, p.33) makes the point that materials play a vital role here and can assist students to

identify, in an authentic way, what they have done.

Rea (1983, p.89) lists possible strategies for evaluation:

(i) measurement;
(ii) self-assessment;
(iii) observation;
(iv) record/diary keeping.

In a later paper (Rea 1987, p.152) she gives us a more elaborate list of means that can be used to derive data. She makes two distinctions, firstly between quantitative and qualitative means and secondly, between formal and informal means. Furthermore she emphasizes that a range of different procedures, as well as input from a variety of sources, should be used in order to obtain reliable and comprehensive information.

The evaluator will have to make the ultimate decision as to what methods will be most suitable, according to the environment. "What works well in one environment may not produce the desired responses in another, different situation" (Rea 1987, p.152). Examples of the means of collecting information that Rea suggests are :

Formal
- systematic review of curriculum materials;
- tests;
- meetings with instructors;
- guided questionnaire;
- structured interviews.

Informal
- generally more open-ended feedback such as reports, unguided and open questionnaires.

A final point to make is that methods suitable for product evaluation will tend to involve measurement, whereas methods that are used for process evaluation will more likely be descriptive (Rea 1983, p.97).

The main conclusion here is that the methods of evaluation chosen need to be appropriate to the situation. It would therefore seem sensible to have a bank of methods from which to choose, according to the desired criteria, according to what is being evaluated, and according to who is evaluating and the state of their knowledge.

2.6 Should evaluation be summative, formative or both?

We have already discussed Rea's (1983) point that different audiences need and expect different kinds of information; we have examined the process/product issues that influence what we evaluate; and we have seen that evaluation may be descriptive and serve an explanatory function or, on the other hand, serve as the basis for decision-making.

Now, another important distinction to make is between evaluation which "serves as a feedback to participants in the evaluation process" and evaluation which "serves as a guide to future developments" (Rea 1983,

p.90). The idea of guiding future development through evaluation is what is commonly referred to as formative assessment, since this part of evaluation forms an integral part of the whole idea of project design development. This allows for continuous development and change to take place within the curriculum.

The other kind of evaluation procedure is summative assessment. While formative evaluation is concerned with process, this is concerned with product. It is likely to be quantitative in form, and to be used largely for decision-making purposes. Examples of this type of evaluation method would be formal means of measurement, such as standardized tests and examinations.

A major distinction which Rea (1983, p.90) makes between the two types of evaluation is that formative evaluation is likely to be more subjective in nature. She mentions methods of data gathering such as self-appraisal, interviews, and observation. This represents a move away from a more traditional approach to gathering facts, towards a new approach which she calls information exchange. The implications of the two types of assessment are different. Formative evaluation implies a built-in allowance and desire for change to a programme or curriculum. We have already seen that Rea recommends flexibility and responsiveness to change and innovation in a programme. Formative evaluation provides the machinery for this.

In her definition of a communicative curriculum, Rea (1987, p.149) emphasizes "learner uptake (i.e. behaviour) as something that cannot be preordained and predicted". She says that in the past test procedures have been too narrowly "prescribed as quantitative and restricted to the measurement of the outcomes (i.e. the products) of the learning process". She redefines evaluation in broader terms, one of which is: "negotiable, arising out of the teaching and learning, that can in part be specified in advance, but will also be open-ended" (Rea 1987, p.150). The implication of this is that while there is a place for summative evaluation this should not be to the exclusion of formative evaluation. We have already discussed different forms of evaluation and they can be seen as valid in terms of the different functions they serve. For example, there are needs created by the internal development of the curriculum, and needs created by public accountability. According to Rea (1987, p.151), both types of evaluation can provide different kinds of evidence of the effectiveness of programmes.

Murphy (1985, p.4) similarly tends to equate evaluation of product, generally learner performance, with summative evaluation, and evaluation of the learning process with formative evaluation. His interest in the whole question is in the consequences for teaching and curriculum development. His view is that formative evaluation, because it is process- rather than product-oriented, is more likely to encourage participants, especially teach-

ers, to get involved in the process. In this way it is more likely that there will be co-operation and openness, rather than professional secrecy. A consequence of this could be a high level of accountability and professional awareness among teachers (Murphy 1985, pp.14 and 16). This is not to deny the usefulness of quantitative and summative evaluation, but Murphy stresses that up to now an over-emphasis on this aspect of evaluation has meant little development in other aspects. And of course he is most concerned with the idea of involving people other than the outsider/expert in the whole process.

In Potts's opinion, the communicative view of language teaching and learning has far-reaching implications for the type of evaluation that needs to be adopted. The main implication is that the communicative curriculum does not have the easily identifiable and quantifiable internal criteria that we find in structural and functional curricula (Potts 1985, p.20). Similarly, while the objective curriculum could be seen as static and not vulnerable to change as a result of assessment, the communicative curriculum must be responsive to evaluation (Potts 1985, p.22). Potts concludes that an accurate procedure for both formative and summative assessment is vital. He holds that the quality of learning in any course in progress depends on formative assessment, while future curriculum development depends on summative assessment (Potts 1985, p.22).

The three writers under review are at pains to stress the importance of formative assessment, mainly because they feel that it has been neglected up to now, because the machinery for it has not been available and the need for it has not been perceived. At the same time there is clearly a need for the continuation of summative assessment.

2.7 At what stages in the programme should evaluation take place?

Traditionally, evaluation is retrospective and takes place at the end of a programme. The direction of our argument so far has been that this should change. With the stress on formative evaluation must come changes in every aspect of evaluation. We therefore need to look again at the question of when in a programme evaluation should take place.

The emphasis on assessment of the process of learning as a basis for change would suggest that evaluation should be happening throughout the programme and not just at the end. Furthermore, Murphy does not see evaluation as distinct from the language teaching process, but as an integral part of it (Murphy 1985, p.16).

In this connexion we might consider the model of the curriculum put forward by Breen and Candlin (1980, cit. Murphy 1985, p.7), in which evaluation is a central component. They argue that since evaluation is a natural part of communication, the evaluation of learners must become a

continuous process within the curriculum. The point here is that whenever there is communication in the classroom there is also evaluation. However, the weakness of their line of argument, according to Murphy (1985, p.8), lies in the lack of systematicity of evaluation provided for in this model. If learners are to evaluate the grammaticality, coherence, appropriateness and intelligibility of utterances in the classroom in a way that provides useful feedback to other learners, they will need to be taught how to do it. Furthermore the question is raised as to whether the kind of evaluation that goes on in real communication is different to the kind of evaluation necessary for classroom learning. Murphy (1985, p.9) concludes that we need to know how such theory works in practice. This means more empirical testing of such ideas in the classroom.

We return now to Potts's idea of evaluation as the "motor of the curriculum" (Potts 1985, p.19). He also sees evaluation as a continuous process, and the constant driving force behind curriculum development. Thus for him too evaluation cannot be relegated to the last stage of a programme. Potts advocates giving students some responsibility for their own evaluation right from the start.

Rea (1983, p.88) proposes a three-tier evaluation system—initial, on-going, and retrospective. It is not sufficient to provide for evaluation when the project has been set up and is running. Instead it should be built in at the design stage and begin prior to the planning stage. Rea (1987, p.150) makes the point that evaluation at different stages in the curriculum process may be undertaken for different reasons. For instance, if evaluation is related to the process of learning and has a pedagogic function, then it will take place during the learning process. On the other hand, if we wish to satisfy external requirements and provide the administration with product figures, then evaluation will be retrospective and will take place after learning.

Rea (1983, p.93) gives an example of an evaluation plan where she shows that external and internal requirements need not conflict. To satisfy internal pedagogic requirements, regular checks of student performance are planned into the course in the form of tasks. At the same time, in order to satisfy university requirements, formal tests are planned in also.

Potts, Murphy and Rea agree that evaluation cannot just be left to the end of a course, but must be planned in right from the start. Some assessment must be on-going and occur at regular stages throughout the course so that its results can be continuously fed back into the course.

2.8 What channels should be used to feed information back into the programme?

Potts (1985, p.33) suggests that "multiple feedback cycles" are necessary in order to capitalize on the intrinsic communicative role of evaluation

within the classroom. The idea is for feedback from evaluation to be conducted within an open communicative environment, and, as we have already said, this must involve the learners directly, under teacher guidance. Evaluation, according to Potts, is a "public concern" and should not be seen just in terms of a teacher/learner concern, but as a whole negotiation within the classroom between learners, between learners and materials, and so on.

Accordingly, we conclude that evaluation feedback is used to inform the learning process and to further foster real communication in the classroom. Evaluation feedback also has a function as far as motivation is concerned. It is now intended to direct future learning activity in so far as it promotes on-going debate in the classroom about what is being learnt and how it is being learnt. According to Potts, this in itself may provide a useful and authentic source of language learning.

These arguments imply that there need to be many channels within the classroom for feedback in all directions and involving all elements. It is not only a question of the teacher channelling feedback to learners. It is also a matter of learners channelling feedback to the teacher, activities and tasks acting as a channel to all concerned, learners channelling feedback to each other, the learning programme channelling information on the learning process to learners, teachers channelling feedback to other teachers or other people involved in the programme, and so on.

We have seen that among Murphy's reasons for evaluation is the idea that we assess whether current theory works in practice. This means that while he assumes that teachers need to take more responsibility for evaluation, the information that is collected must eventually be channelled to those involved in formulating the theory on which curriculum development is based.

For Rea (1983, p.88) proper discussion and adjustment of the design document is a necessary prerequisite for the smooth running of the project. This means that procedures for evaluation should be planned into the design document. At the same time, the feedback from evaluation will inform improvements and adjustments to the document, both in relation to the current course and with future courses in mind. Consequently there need to be channels which allow this information on the running of the programme to filter back to the designers.

Although there are differences in emphasis between these three views on what should happen with evaluation information, they all consider that a project, course or curriculum needs to be well-informed about the realization of the plan. This implies that it is necessary to create reliable and efficient channels for the dissemination of information, wherever it is needed.

2.9 What kind of constraints can we expect to encounter in the evaluation of a programme?

It is generally recognized that the evaluation of a communicative curriculum is problematic. The lack of an inventory of items to be learned means that the assessment of a programme is less straightforward and therefore more difficult to operate. This helps to explain why there is a lack of well worked-out models of programme design for the communicative curriculum that include evaluation, especially of the formative kind.

In fact, judging from the literature available, there has up to now been relatively little interest in or investigation of the whole area of on-going evaluation. While there are many papers on "evaluation", for the majority of writers this means testing student performance in a more or less summative way. Indeed, it seems that what is called "on-going, formative" evaluation is not always what it is meant to be. As Potts (1985, p.25) points out, if on-going assessment does not influence subsequent activity and syllabus content, then it is in fact not a formative assessment, but a series of "micro-summative" evaluations.

Given that the planners of a project agree that an on-going evaluation plan needs to be built in, there are still difficulties to be faced. Firstly, the capacity of those involved in the evaluation needs to be taken into account. A major problem here will be that teachers, learners, course designers, and administrators will be used to implement summative evaluation procedures and may not have the training to plan or carry out formative evaluation. A further constraint might be time—the time required for gathering and analysing data, and the time required to use the information gathered in order to shape further development of the course in progress. Allied to this point is the amount of effort and resources that it is practicable to expend on this aspect of course planning. Another constraint concerns the channelling of information between participants in the evaluation. This may have many causes:

- physical distance;
- personality clashes and conflict within the hierarchy;
- inadequate channels of communication;
- traditional role relationships within the wider educational environment.

Any one of these, or a combination of them, could make the kind of evaluation procedure we have been discussing here very difficult to implement.

2.10 Conclusion

The three writers on whom we have based our study in this section were chosen because of their commitment to the implementation of procedures

for formative evaluation. They have provided us with a thorough exposition of the reasons for on-going evaluation and of issues that have to be taken into account if this kind of evaluation is to be applied to project design. We are now in a position to summarize the main features of a comprehensive system of evaluation which can play an effective role in programme planning and implementation:

(i) The evaluation needs to be broad enough in its scope to be able to validate the programme design and the theoretical concepts behind it, as well as to assess the success of the programme and of the learners.

(ii) It needs to provide a focus on process as well as on product.

(iii) It needs to be based on clear criteria which all participants are aware of and which evaluators can apply in analysis.

(iv) It needs to involve all participants, non-experts as well as experts, in programme design and modification.

(v) It needs to specify methods of data collection which are appropriate to the context, to what is being evaluated, and to whoever is evaluating. These methods must be reliable and accessible to all concerned.

(vi) It needs to provide suitable and reliable channels to feed the information collected back into the planning of the programme.

(vii) It needs to include formative as well as summative procedures, to be planned into the programme from the start and to be carried out at regular intervals during and after the course.

(viii) It needs to take into account all potential constraints that might arise in the context of a programme, and the design and evaluation of the programme should be adjusted accordingly.

If we attempt to incorporate these features into a project, the task of providing for evaluation will become a much more rigorous and demanding one than we have hitherto known. We must therefore go on to ask ourselves if this effort would really pay dividends in the form of improvements to our projects. Do we really want to expend so much time, energy, and resources on developing alternative procedures for evaluation? Would the data we collect enable the programme to become more efficient or effective than present programmes, which depend on well-designed tests and aware and interested teachers?

3 EVALUATION AND INNOVATION

So far in this paper we have looked at project design from various perspectives. In the first section we examined a variety of project types to see what elements of evaluation they contained, and in the second section

we looked at some theories of evaluation in project design. Now we consider another aspect of project design, namely the extent to which projects introduce innovation. We begin by examining different features of innovation in projects and discuss the extent to which planning of change is necessary in project design. Then we go on to examine the contribution that the type of evaluation we described at the end of the previous section could make to the planning of innovation.

For ideas on innovation in project design we turn to Kennedy (1988). The starting point of his argument is that all English language teaching projects imply change to a greater or lesser degree. Thus he argues that in planning and managing projects skills are needed for this aspect of programme design. He stresses the need to manage change in the initial stages of a project, during project implementation, and on completion (Kennedy 1988, pp.329f.).

We can already establish a link here with the type of evaluation that is suggested by our discussion so far. Therefore a further reason why we should plan for on-going evaluation might be in order to manage change. However, if we are going to adopt Kennedy's argument as a justification for on-going evaluation, we need first to clarify what we mean by innovation in English language teaching projects and then to plan how evaluative procedures might be used to manage change

3.1 Innovation and the project

a) Plans to introduce innovation

A project is essentially a means by which we introduce innovation into a system. If we look back over the projects which we examined in section 1, we see examples of desired changes, either in the form of the introduction of something entirely new (the Study Skills in English project and the Somali textbook project) or by modifying something that already exists (the Sierra Leone project, the Ain Shams project).

These projects and the changes they bring with them proceed in defined stages. Kennedy (1988, p.329) broadly categorizes these stages as:

- identification of a problem;
- selection of the innovation;
- incorporation, acceptance and diffusion of innovation.

An example of this can be seen in the Somali textbook project. The need to develop new materials was identified by the Somali ministry, the innovation was decided on and worked out by KELT officers, and the resulting new materials and changes in teaching methods were introduced gradually and systematically as part of the programme. Kennedy (1988, p.329) makes the point that "changes brought about will not only be quantitative, they will also be qualitative, they will involve people as well

as things". Therefore, in the case of the Somali project, it was not only the organization of material resources that needed to be managed, but a great deal of sensitivity was necessary in the management of the changes that resulted.

This kind of sensitivity was also stressed by Bowers when describing the Ain Shams project. He too made the point that any curriculum development project is an attempt to introduce change "in a common set of interlocking professional and administrative systems" (Bowers 1983, p.100). A change at any point in the system entails some change at every point. Bowers (1983, pp.100ff.) describes the set of systems which curriculum innovation affects and defines the issues which arise in each case. He comes up with a "spider's web" of interlocking systems which include personnel, facilities, funding and management. Kennedy (1988, p.331) adds social, political and cultural systems to this list. "Whenever the spider's web is touched, the whole trembles" (Bowers 1983, p.104). Bowers points to the need for awareness of the effect of this interdependence on the project as a whole.

b) Planning for on-going innovation

Change is clearly a feature of the life of a project, and this must be taken account of. If we accept the views of Kennedy and Bowers, then reliance on an *ad hoc* system that depends on the sensitivity or intuitions of planners and administrators is not sufficient. We need a project plan that:

- "can be monitored and adjusted as the process of change is taking place" (Kennedy 1988, p. 330);
- can look back over a project and relate the process to the outcome in order to investigate reasons for successes and failures;
- can take this information forward to future project development.

An example of such a project plan is the Tanzanian syllabus and materials design programme described in section 1 (Brumfit 1980). The course was structured in such a way that change was not only allowed for but also encouraged. The framework for the course was not a fixed scheme, but was based on a syllabus and a set of principles that acted as a working document. This meant that there was capacity for change during the running of the course. In the long term this could mean the replacement of the original document, a restructuring of the syllabus, and major changes to the teacher-training programme.

c) Systems affecting and affected by innovation

Planning a project means planning for innovation. We have already seen that there are certain features of innovation that we need to keep in mind. Kennedy (1988, pp. 330f.) highlights some of these features:

(i) the reciprocal effect of innovation on systems and systems on

innovation;

(ii) the importance of innovations being acceptable to local conditions and cultural norms;

(iii) the need to involve and consult all participants in the innovation process;

(iv) the need to show that the individuals affected by the innovation will receive benefits as a result.

Given that a project will affect existing systems, we must also expect that these systems will exercise some influence over the project and the innovation. Kennedy (1988,p.332) offers a plan of a hierarchy of systems in terms of relative influence:

cultural
|
political
|
administrative
|
educational
|
institutional
|
classroom innovation

The point here is that problems will arise

- if the project and innovation management are ignorant of the higher order factors;
- if they do not work within these factors;
- if they try to change them.

This therefore returns us to the point that we made earlier: we need to manage change in the initial stages of the project. But now we go a step further and advocate project pre-planning, at which stage the innovations, their effects on systems, and the way they might in turn be affected by existing systems, will need to be predicted in advance, and the whole context evaluated, before change is ever made. This then needs to be controlled throughout the whole project.

d) Conditions necessary for innovations to succeed and be accepted

(i) A variety of participants will be affected by the innovation introduced by a project. The important question is whether there is communication between the participants, and if so, whether that communication is one-way or two-way.

(ii) All those involved in the project need to agree that innovation is necessary. For example, the teachers need to feel that there is a close match between their working context and the innovation at three levels:
- feasibility in terms of resources and organisation;
- acceptability in terms of the teaching and learning style advocated by the project, and their own philosophy;
- relevance in terms of the needs of the learners.

(iii) Innovation needs to be incremental rather than sudden, so that all aspects of the programme can be adopted.

(iv) "Ownership" is a term Kennedy uses to describe "the degree to which the participants feel that the innovation 'belongs' to them" (Kennedy 1988, p.338). This, he claims, has a great influence over the likelihood of innovation establishing itself.

(v) It is important that teachers feel they have something to gain from innovation resulting from a project, and that these gains outweigh the losses. Kennedy (1988, pp.339-340) proposes a series of possible gains:
- increased job security;
- improved relations with the administration;
- improved service to learners;
- improvements to teachers' knowledge and skills;
- intellectual satisfaction;
- economic and professional rewards.

These need to be off-set against such losses as:
- the time-consuming nature of innovation;
- the pressure on teachers to learn new skills;
- the physical and mental resources required.

3.2 Evaluation as the manager of innovation

a) Plans to introduce evaluation

The decision to introduce innovation into a system necessarily involves introducing some form of evaluation. This is the only means of measuring the effects of innovation as it occurs, making judgements about its success or failure, assessing its repercussions on other systems, and consequently adjusting or continuing with the innovation plan. Accordingly, on-going evaluation needs to be introduced into a project plan in order to manage change from the outset.

b) Planning for on-going evaluation

Just as *ad hoc* planning for innovation is not sufficient, so we need a systematic approach to planning evaluation procedure. Just as innovation

requires a way of relating the process of change to its outcome, so we need to be able to relate the evaluation of the process (formative) to the evaluation of the outcome (summative). If we wish our innovation to be carried on in the system then this relationship is vital.

By providing for on-going evaluation in a programme plan, we are declaring from the outset a willingness to accept change, and to set up the right framework and the right atmosphere for it to occur in. Eventually this may result in a programme plan that is very different from the initial design document.

c) Systems affecting and affected by evaluation

A knowledge of the systems within which the project is being constructed is clearly necessary. But it cannot be expected that the project designer will have direct access to all relevant information. This can be overcome by designing an evaluation system that allows input from all levels of the system. However, we cannot assume that getting relevant information at any particular stage of the programme, for example at the outset, will be sufficient. This is because the repercussions of innovation will continue throughout the programme. The evaluation system has to cope with this and provide continuous feedback on the way in which systems are reacting.

d) Conditions necessary for participants to co-operate in the evaluation procedure

(i) Our evaluation needs to be able to tell us how all participants are reacting to the project. In the top-down hierarchy, for example, there may be communication downwards to the teachers, but there may be no existing channels to allow information to filter back up from the classroom. The evaluator needs to be aware of this and to provide suitable channels. That may mean, for example, that it falls to the course designers themselves to distribute questionnaires and other means of gathering information to the teachers.

(ii) If teachers have a hand in the decision-making process concerning innovations and are asked through the evaluation process for their opinions, then they are more likely to welcome the innovation. It may happen that teachers still do not welcome the innovation, for a number of reasons. One major reason might be that the innovation is at odds with their own philosophy of teaching. Through effective evaluation it might be possible to identify this problem and as a result make necessary adjustments, for example to the teacher-training programme.

(iii) As we have seen from Bowers's "spider's web", a change can bring

about a number of associated changes. An evaluation would be able to assess this ripple effect at regular intervals. This might involve introducing the innovation in stages and evaluating its effects at each stage throughout the programme.

(iv) The idea that teachers make gains or losses as a result of innovation is very relevant to the evaluation of a project. Clearly, the evaluation procedure must be able to assess what burden it puts on participants. This burden can be increased or reduced according to the gains that participants may make. For this reason evaluation will have to find out if there are gains to be made. If there are not, it may be that the designer can promote some. For example, in the Somali project, the creation of teacher associations fostered solidarity among teachers. On the other hand, it may be impossible, because of existing systems, to create gains. In that case, the designer could aim at a reduction of losses. In any case, a project designer should be able to evaluate in advance what gains could be anticipated for participants, and set the targets accordingly.

3.3 Conclusion

To the question: Is it worth taking a new look at evaluation procedures? the answer must be "yes". If we agree that projects should be set up, curricula instituted, materials written, teacher-training courses planned, and language courses implemented, all with the aim of changing something; and if we agree that it is desirable and useful to monitor and shape this change as it occurs; then we must have a suitable way of evaluating the change. There is a great deal in common between managing innovation (in Kennedy's 1988 terms) and conducting evaluation as it is envisaged in this paper.

4 THE MALAYSIAN STUDY SKILLS PROJECT: TEACHER EVALUATORS

Up to now we have been mainly concerned with assessment from the perspective of the course designer, who in most cases is an expert outsider. However, we have also made the point that evaluation can be conducted from the perspective of any of the participants in a project. To continue our inquiry into project evaluation we look at a project in which I was involved. It does not seem sensible to examine it from the course designer's perspective, since I could expect to have only a partial view of this. Instead I will examine the project from my perspective as a teacher implementing the project design in the classroom.

There is a certain difficulty in writing about something that one has been

personally involved in. On the one hand, there is a desire to avoid being critical of the work of former colleagues; on the other hand, there is the the tendency to look back with the benefit of hindsight and suggest how things should have been done. However, even with these defects, it is perhaps useful to look at a project from the point of view of participants at the chalkface and give a different perspective from the umbrella view of designers.

Our purpose is to examine the framework of the project in order to identify and label the elements of evaluation that were built in or included later. To do this we will review the design document and the project in practice, in the light of evaluation features which we consider essential to project implementation, and which we listed at the end of section 2. But first we offer some background information on the project.

4.1 Study skills in English for A Level students in Malaysian residential schools

This project was concerned with Malaysian students who had completed their secondary education and were now studying for British A Levels in preparation for entrance to British universities. Apart from their A Level studies, the students also followed a course in study skills in English (SSE) over the two years. This was a major component, taking up twelve hours per week and designed:

- to facilitate A Level studies;
- to prepare students for British university life;
- to prepare students for a proficiency test in English required by British universities.

The plan was to set up ten centres around Malaysia, in state residential secondary schools, and to admit approximately sixty Lower Sixth students to each centre each year. The Ministry of Education oversaw the project and funding came from the Malaysian Public Works Department (JPA), which in due course would finance students at university.

A course was designed and implemented by a British organisation, the Centre for British Teachers Ltd (CFBT), called in for this specific purpose. The CFBT provided a substantial number of the project personnel:

- a project planner to produce the course syllabus and subsequently to act as professional co-ordinator;
- a project director to perform administrative functions and liaise with relevant bodies;
- regional project directors to supervise the smooth running of the operation in the ten centres around the country and monitor the implementation of the course in classrooms;
- centre co-ordinators who were teachers, one per centre, working

within the schools and liaising between school administration, the CFBT head office, and the teachers;

- teachers—all the SSE staff and approximately half the subject teachers on the A Level programme were CFBT expatriates.

The administration in the individual schools retained a level of control over the A Level programme. The head teacher at each centre was the immediate superior of all teachers, including the expatriate staff. There was an A Level committee of all A Level teachers, expatriate and Malaysian. There was an A Level co-ordinator, a Malaysian colleague appointed by and answerable to the head teacher. The A Level co-ordinator had responsibility for decisions made at each centre and for liaising with the Ministry and the JPA on behalf of his/her committee.

4.2 Assessing the evaluation policy of the project

We shall look at each of the eight main features of evaluation that we listed at the end of section 2 and assess to what extent this project incorporated them.

(i) The evaluation needs to be broad enough in its scope to be able to validate the programme design and the theoretical concepts behind it, as well as to assess the success of the programme and of the learners.

The design document allowed for assessment of student progress and course success. In Floyd (1986, p.60) task-based assessment is defined as follows: "Progress towards each course objective can be tested by setting specific pieces of work [...]. These tasks can be graded in difficulty and assigned a value on a band scale [...]. Alternatively, criteria can be established for performance at each band level on the same task." This plan for graded tasks allows us to judge the progress of the course and the performance of particular students in terms of product. If used in all centres, it would also allow evaluation of the relative progress in the different schools, which might be useful for external evaluation requirements. This is probably the type of assessment that Potts (1985) would label "micro-summative" in that continuous assessment of student performance is provided for, but the focus of the evaluation is summative rather than formative. The criterion used for judging success is the extent to which students, and by extension the course, meet course objectives.

Clearly, this document is based on theoretical concepts. While they are not explicitly discussed in the document, they are evident in its contents. For example, it is taken for granted that participants will agree that such concepts as independent learning, task-based learning, integrated skills work, student self-evaluation, and banding form the basis for an effective approach. We have already argued that an evaluation procedure should be

sensitive to the possibility that the theoretical concepts of the document might be in conflict with participants' philosophies. Therefore some means of evaluating this match or mismatch needs to be present so that in the case of conflict an adjustment can be made to the document or changes made to teacher recruitment/training.

As far as validation of the thinking behind the document is concerned, the document provides us with an opportunity to get information on whether the approaches to learning and teaching inherent in the project are working (Floyd 1986, pp.62-3). The most obvious way of getting such information is probably student self-assessment, with feedback to SSE teachers and eventual improvement of the course as the goal.

In practice the project followed the design document closely as far as the monitoring of student progress was concerned. As regards evaluation of the theory behind the course, channels were planned in to gather teacher and student reaction throughout the running of the course. For example, at the beginning of the project and at each new intake of teachers, a two-week orientation course provided an introduction to the thinking and method behind the course in workshop sessions. Subsequently, there were a number of possible channels for teachers to feed back their observations on approaches and methods being used on the course:

- regular seminars and workshops based on contributions from teachers and attended on a voluntary basis by teachers from all centres and by CFBT administrative staff;
- regular visits to the centres by CFBT regional project directors (RPDs)—during these visits the opportunity was given for discussion of classroom methods and materials, both on an individual and group basis and RPDs observed teachers in the classroom and wrote reports in the form of lesson descriptions;
- a regular internal publication, *A Level Exchange*, which contained lesson ideas, views on different approaches, and so on;
- termly meetings between co-ordinators from all the centres and CFBT staff to discuss, among other things, the progress of the course—at these meetings co-ordinators represented the views of the teachers at their centres, discussed and agreed subsequent changes to the programme, and carried news of these changes back to the centres;
- meetings with teachers at the centres followed to discuss changes—reports of these meetings were returned to the CFBT, where disagreements were noted;
- from the point of view of one teacher at a particular centre, a very fruitful channel for the evaluation of approaches and the best source of feedback was the closely-knit SSE department at that centre.

At a formal level the channel of communication was weekly meetings of the six-strong department. This was a means of filtering out ideas to be communicated to the CFBT hierarchy through the co-ordinators' meetings and disseminated *via* the internal publication. However, it was really at classroom and staffroom level, through team lesson planning and team teaching, that the most useful evaluation of basic approaches took place. For various reasons, the more formal channels listed above did not always work as they were intended to. For example, attendance at workshops was haphazard; and there is no doubt that distances between centres made communication difficult. For the same reason, RPD visits were not as frequent as they might have been and often concentrated more on welfare concerns than classroom matters. Also, a large part of the centre co-ordinator's time was spent on bureaucratic details which limited the time available for evaluating the progress of the course.

(ii) Evaluation needs to provide a focus on process as well as on product.

The design document accords great importance to the achievement of good A Level results and to the English proficiency exams. The target product of the course is a student who has acceptable levels in his academic subjects and in English, and who as a consequence gets a place at university in Britain. However, to achieve this product, attention to the learning process is vital and this is reflected in the SSE course objectives:

"1. To develop the language, thinking and independent learning skills needed to achieve the overall aims [i.e. success in A Levels and university entrance].

"2. To facilitate the transfer and application of these skills to A Level subject learning" (Floyd 1986, p.1).

The opportunity to evaluate the evolving learning process is built in through the self-assessment component in the form of self-assessment forms completed by students, learning diaries and tutorials. It is also allowed for in the continuous assessment plan through feedback from subject teachers and possibly through tasks to assess needs and progress, if these tasks are focussed on the process rather than on the product. However, the machinery for doing this was not built into the design document, and individual teachers were left to devise it. For instance, the document gives no indication of what a self-assessment form might look like; no criteria are given for evaluating information gathered from learners in diaries or on self-assessment forms; and no procedure is suggested for eliciting information from subject teachers helpful to the evaluation of process. This prompts us to make a general comment about the focus of the evaluation within the document. There is a heavier emphasis on monitoring and evaluating performance; and while some attention is paid to learning

processes, there is far more focus on the evaluation of product.

In practice the project proved that the design document assumed a lot on the part of the subject and SSE teachers involved. In fact, very little systematic evaluation of the learning process took place at the outset of the course because of the lack of instruments and guidance. However, as time went on, within individual centres the very diverse aims of the two directions of the course (i.e. A Level and SSE) began to merge as teachers worked together, discussed mutual problems, and sought solutions that were compatible with both sets of aims. Thus it emerged that joint attention to the process of learning and teaching was the most effective and consistent way of yielding an improved product, i.e. better English and subject results.

(iii) Evaluation needs to be based on clear criteria which all participants are aware of and which evaluators can apply in analysis.

The design document contains guidelines as to what is expected of student performance. These were expressed in band scales (Floyd 1986, pp.66-70), were widely circulated, and became familiar to students.

We have already established that the design document advocates a task-based approach. We would therefore expect to have at our disposal a range of criteria by which we could judge whether our students were able to perform the tasks, not just in terms of product, but also in terms of the process involved in performing that task. However, the document does not provide this in a very detailed way (Floyd 1986, p.26). The idea is that the teacher should be able to use these criteria as a checklist to note both successes and failures and feed that information into subsequent lesson planning. One comment we need to make here is while we have guidance for the collection of the evaluation data, there is no guidance on how to make use of it. In practice it proved straightforward to rely on the design document for performance criteria. For tests, the band scales came to be replaced by closely related ones, namely those of the English Language Testing Service exam.

As far as the introduction of new learning processes was concerned, the teachers had to evolve their own criteria for evaluation. These were based on teachers' own experience of the learning context in Britain and elsewhere, as well as observations by teachers of existing learning processes being applied to A Level materials in Malaysian classrooms. Teachers quickly observed that there was a great gulf between the learning processes required by A Level and university study and those that students were actually applying. The latter were clearly insufficient. The criteria therefore were based on the teachers' own judgement of what processes were necessary and they observed learners to assess such things as:

- whether they adopted a problem-solving approach;

- whether they tended to learn by rote;
- whether they were capable of tackling problems in groups;
- whether they could split up tasks into relevant sub-tasks;
- whether they were able to transfer skills from SSE to A Level.

(iv) The evaluation needs to involve all participants, non-experts as well as experts, in programme design and modification.

The design document acknowledges that teachers contributed to it by submitting schemes of work, suggesting teaching ideas, and offering examples of materials. It goes on to say: "[...] to a large extent the syllabus is a reflection, now in booklet form, of the teaching which is already going on in the various centres" (Floyd 1986)

This is an example of a design document which appeared after the implementation of the programme. The development and the implementation of the programme started around the same time. Bachman and Strick (1981, p.51) comment that this represents a typical sequencing of a new programme in which there is little or no time for development before implementation. The advantage of this is that there is time to involve participants in the development, so that some evaluation of the document by participants is possible.

The completed document contains suggestions for a teaching programme, but imposes no rigid structure. Therefore, in so far as teachers were able to use it as a framework, they designed and modified their own courses according to the feedback they obtained in their own classrooms.

The design document indicates that subject teachers should also be involved in SSE programme design in so far as they are able to evaluate the current needs of students (Floyd 1986, pp.8f.). In practice this proved possible, but it took a long time to develop suitable means by which to elicit from subject teachers what exactly student needs were, in terms that SSE teachers could translate into study skills. What was lacking was an evaluation instrument the language of which was accessible to both SSE and subject staff. This kind of co-involvement of subject and SSE staff in identifying learner needs varied from centre to centre and depended to a large degree on the co-operation that existed between teachers.

(v) The evaluation needs to specify methods of data collection which are appropriate to the context, to what is being evaluated and who is evaluating. These methods must be reliable and accessible to all concerned.

The design document refers to a number of different means of data collection under self-assessment, peer assessment, continuous assessment, common tests, and external exams (Floyd, pp.62f.). All of these are concerned with the evaluation of students' performance only.

In practice, methods of data collection also included introspection on the

part of teachers and observation of teacher performance. This developed over time, but not on any formal basis. The methods of data collection which were used most often were:

- informal teacher observation of own lessons;
- joint observation during team-teaching and subsequent discussion;
- observation of the measure of success/failure of lessons planned by other teachers and subsequent discussion;
- RPD observation of individual lessons and feedback to the teacher concerned in the form of a report;
- discussion with students on a tutorial basis about learning progress;
- elicitation of feedback from students on their language needs, the appropriacy of the content of the course to their needs, and their estimation of their own progress;
- standardized tests—including English placement tests used for streaming, tests in English and A Level subjects common to all A Level centres, mock O Level and ELTS tests of English, actual O level and ELTS tests.

(vi) Evaluation needs to include formative as well as summative procedures, to be planned into the programme from the start and carried out at regular intervals during and after the course.

The design document, as we have already seen, details types of summative evaluation of student performance and suggests that formative evaluation procedures would be a good idea, but it does not detail them. The main evaluation that the document dwells on is the test performance of students and this is summative.

In practice, formative evaluation played a vital role in the programme. As we have seen, individual teachers, or teams of teachers, at the different centres had the task of designing their own teaching programmes on the basis of the loose framework of the design document. They were therefore highly reliant on their own evaluation of classroom performance. The success of methods and techniques reinforced existing approaches and failure led to a search for alternatives and modification of the teaching programme. So, in effect, the looseness of the teaching syllabus encouraged this type of formative evaluation. It should be noted that the project recruited only qualified and experienced teachers to work on the SSE programme. We might conclude that this is a prerequisite if this type of evaluation is to be realized in this type of framework. Furthermore, the support provided by the organization in the form of professional development could also be considered significant.

(vii) The evaluation needs to provide suitable and reliable channels to feed the information collected back into the planning of the programme.

The design document does not distinguish clearly between methods for collecting data and channels for feeding this information back into the programme. As we have seen, there are suggestions for the gathering of information, but there is no detailing of means for analysing and re-utilizing this. This may be because the methods that are suggested are mainly carried out by teachers who have a direct involvement in the implementation of the programme. Therefore it is perhaps not felt necessary to specify channels. The channel that the programme relies heavily on is the three-way communication process between subject teachers, SSE teachers and students.

In practice the problem was that good communication tended to rely to some degree on good personal relations between teachers, rather than being provided for in a formal way in a programme. This can be seen in Floyd (1986, p.48).

The last stage of the flowchart indicates that evaluation should take place, but precisely what should be done with the information is not specified. Ways in which this channel actually operated included:

- the subject teacher observing students performing a subject-related task in the SSE classroom to allow him/her to note recurring problems with subject material and incorporate this into his/her programme;
- the subject teacher observing students successfully performing activities in the SSE classroom, using techniques that he/she could adopt for the subject classroom (e.g. a group approach to problem-solving which was largely developed in the SSE classroom and carried over to subject teaching/learning through this channel);
- the SSE teacher observing students in the subject classroom and noting language and study skills problems which were subsequently dealt with in SSE lessons;
- the marking by subject teachers of subject-related assignments set in SSE lessons and subsequent feedback into subject and SSE programmes;
- the provision for one-to-one or group dialogues between students and teachers to discuss current problems and possible solutions.

(viii) Evaluation needs to take into account all potential constraints that might arise in the context of a particular programme, and adjust the design of the programme and its evaluation accordingly.

I cannot comment on the constraints that shaped the design document since I was not involved in that aspect of the programme. Instead, my comments will be confined to the constraints that were observed in practice and how these shaped the evaluation procedures used.

The first constraint that we experienced was a variation among the

participants in their capacity to put this type of evaluation into practice. Allied to this was a great diversity in teaching philosophies, and without wishing to oversimplify, this generally divided SSE from subject teachers. In some cases this remained an obstacle throughout the project and meant that summative evaluation remained dominant in certain areas of the programme. In other cases there was a coming together of approaches which eventually meant that more formative evaluation was possible. This was facilitated by a professional development programme which allowed teachers, both subject and SSE, the opportunity to discuss their ideas and methods. Related to this was the need to involve students as participants in self-evaluation. For cultural and educational reasons this had to be a slow process. The solution was to take an incremental approach and introduce the method gradually.

Another constraint was time, in the sense that gathering information could be a lengthy process. Allied to this are the effort and resources involved in collecting the data and re-planning the course accordingly. Our observation was that if changes came down through the hierarchy, then the constraints of time and effort could be obstacles to the evaluation process. However, for decisions taken at centre level, the same machinery was not necessary, the constraints were not as serious, and the changes to the programme were relatively rapid and effective.

The one disadvantage of decision-taking at a local level was perhaps a lack of systematicity. Furthermore, there was a tendency to "re-invent the wheel" at each of the centres, simply because results of evaluation were not filtering back to the hierarchy and being disseminated to the centres in a usable form. Very often, information that had travelled these channels was so idiosyncratic and personalized that it was comprehensible and relevant only to those from whom it originated.

It could be argued that a major reason why time at a local level was not a serious constraint was because of what Kennedy (1988, p.338) calls the loss and gain calculation. There is no doubt that teachers involved in evaluation were expending time and energy over and above the demands of their contracts. The system relied to a certain extent on the personal satisfaction that teachers obtained from this type of involvement. But it also provided "gains" to off-set "loss" through opportunities for professional development and other incentives to co-operative working and seminar participation.

We have already mentioned that the loose framework of the design document relied to a great degree on co-operation between teachers at the centres. While on the one hand the successful working of these relationships could provide the "motor" for effective on-going evaluation, breakdown in the relationships could be detrimental to this aspect of the programme. This

meant that such factors as personality, experience and teaching philosophy played a very important role.

4.3 Conclusion

We can see that the Malaysian Study Skills project, with its comprehensive design document, does contain a number of elements of the type of evaluation we would see as important. At the same time there is a need to translate these ideas and plans into a format which is readily usable by the participants, written in a language that is common to all, and which provides simple methods of analysis to make the best possible use of the data.

5 GUIDELINES FOR A PRACTICAL APPROACH TO ON-GOING EVALUATION

Drawing on our understanding of what on-going evaluation should ideally involve, considering the theoretical implications of implementing such a programme, taking into account that a driving force behind evaluation should be the management of change, and finally having looked at an example of a working system of evaluation, we can now attempt to draw up a set of guidelines for applying on-going evaluation procedures. Our aim here is to make suggestions that would bring the system of evaluation described in section 4 more into line with the ideas explored in the previous sections. At the same time we intend that the suggestions made should be appropriate to a broad range of situations. An important consideration in formulating these guidelines is to put forward ideas which are based on the theory that we have discussed, but which are at the same time immediately applicable in real contexts.

Guideline 1

First and foremost, what we are concerned with here is the production of a design document which offers as much support as possible for methods and techniques of evaluation since:

- the design document and syllabus is the one element of a project which has the potential to reach all participants;
- it can transcend problems posed by poor channels of communication such as geographic distance between centres or personality clashes within a centre;
- it is a starting point for all discussions about the programme;
- it can be a prop when participants need ideas on methods and techniques;
- it can be a launching pad for further developments;

- it is a source of continuity when personnel at various levels change or ideologies are thrown into question.

For all these reasons the design document is very important. Therefore everything which is important to the project should be included, explained, and/or exemplified in it. Since we feel that evaluation is one of the important elements in any project, it too needs to be explained as clearly and fully as possible in the design document. For this reason the suggestions that follow are all related to the design document.

Guideline 2

The design document should make clear what teaching philosophy lies behind the programme and what this might involve in terms of approaches to teaching and learning, and the selection and use of methodologies and materials. By extension, this will involve making explicit the approach to evaluation that the document takes.

It is important for the evaluation of the programme that participants should be aware of the thinking behind the design in order to be able to contribute to the evaluation of that design. If a teacher does not understand why a particular type of activity is present in a programme, there is every possibility that he/she will fail to use this activity correctly and consequently will not be able to judge whether the activity was successful or not. Furthermore, the objectives of a course will be closely linked to the teaching/learning philosophy behind it. If the success of our course is measured according to the extent to which learners meet the objectives, then we must rely heavily on all participants understanding what meeting the objectives involves. And to do that they need to understand the theoretical concepts behind them.

To fulfil these requirements the design document should:

- include an explanation of the theory, and of associated approaches and activities;
- express that explanation in a language and format accessible to all participants, expert and non-expert alike;
- make clear that the principles set out in the document are not fixed but are open to change;
- list suggested techniques for evaluation.

Guideline 3

The document must provide participants with the language they need to collect, understand, and analyse data relevant to evaluation. Unless there is a common language there will always be the problem that data collected by different participants will vary and will only be useful to those who collected it. Consequently it will not be easy to analyse the available data

centrally and disseminate it for wider use. Means of providing a common language might be:

- to prescribe precise methods of data collection;
- to provide examples of methods of data collection which are reproducible and usable as they stand, e.g. questionnaires, checklists, assessment forms;
- to recommend the submission of certain data to the managing body at regular intervals.

Guideline 4

The document needs to provide criteria for the evaluation of process which are just as explicit as those for the evaluation of product. These criteria are not only the domain of experts (course designers, administrators, English or SSE teachers), but should be available to all participants, including students and subject teachers. The design document should give lists of criteria for the evaluation of process that can be reproduced for students and teachers. These can then serve as a basis for discussion in the light of the experience of participants during the programme.

Guideline 5

The design document needs to provide participants with clear ideas on how collected data might be analysed. For this participants need to understand the value of certain types of data and how much information this data can give about the programme. In order to do this, the document could:

- contain worked examples of data and their analysis;
- offer ideas on what particular results might signify for a programme;
- suggest what remedial action might be taken based on particular results;
- recommend how results can act as a basis for discussion of and change to a programme;
- show how student access to these results can act as a useful basis for individualized learning and self-assessment.

Guideline 6

The design document needs to list, describe and explain the implications of all the possible channels that could be used to feed information back into the programme. This will mean:

- giving precise instructions on how to use the various channels, for example prescribing how often teachers might meet;
- making explicit the aims of using certain channels, especially in cases where participants might be mistrustful (for example, teachers should

have access to examples of lesson observation sheets, which could be a useful channel for disseminating good lesson descriptions and ideas);
- making explicit the overall plan for the co-ordination of channels;
- making explicit the way that formal channels work—for example, explaining how Ministry decisions are incorporated into the programme, or how information from teachers is processed.

Guideline 7

The design document needs to include a fully worked out and clearly explained process for the systematic redevelopment of the programme and, by extension, of the design document itself, based on information from the evaluation system. This would involve:
- specifying a timetable for staged reassessment of the programme;
- specifying a procedure for the collation of the data from the whole evaluation machine;
- specifying how the consequent changes to the programme will be decided on.

Guideline 8

The design document must make clear that this approach to evaluation is an incremental one. What happens at the beginning of the programme might not necessarily be what happens at the end, and changes will depend on the system of evaluation to which all participants may make a contribution. In other words, although the document prescribes procedures, this does not mean that they will necessarily always remain in place. But at the same time the document must make clear that changes to these procedures should be discussed and justified.

As a result of these suggestions, the design document would inevitably be a larger document than is usually the case; and its role would be more substantial than that of the usual course syllabus. Indeed, the document we propose here might be described as a "course companion". Certainly such a design document would become a vital tool in the planning, management, implementation and evaluation of the course which was based on it.

REFERENCES

Alderson, J. C. (ed), 1985: *Evaluation*. Lancaster Practical Papers in English Language Education 6. Oxford: Pergamon, and Lancaster University

Bachman, L. F., and G. J. Strick, 1981: "An analytic approach to language program design". In R. Mackay and J. D. Palmer (eds), *Languages for Specific Purposes: Program Design and Evaluation*. Rowley, Mass.: Newbury House

Bowers, R., 1983: "Project planning and performance". In Brumfit, pp.99-120.

Breen, M. P., and C. N. Candlin, 1980: "The essentials of a communicative curriculum in language teaching", *Applied Linguistics* 1.2, pp.

Brown, P. J., and S. B. Hirst, 1983: "Writing reading courses: the interrelationship of theory and practice". In Brumfit, pp.135-150.

Brumfit, C. J., 1980: "Education, ideology, and materials design: a Tanzanian experience". In Smyth, pp.164-171.

Brumfit, C. J. (ed.), 1983: *Language Teaching Projects for the Third World*. ELT Documents 116. Oxford: Pergamon, and The British Council.

Candlin, C. N., J. M. Kirkwood, and H. M. Moore, 1977: "Study skills in English: theoretical issues and practical problems". In Mackay and Mountford, pp.190-219

Clarke, D., H. Hawkes, N. Pritchard, and B. Smith, 1983: "The English Language Textbook Project, Somalia". In Brumfit, pp.49-68.

Floyd, J. L., 1986: *Study skills in English for A Level students in Malaysian residential schools: course syllabus*. Petaling Jaya, Malaysia: Centre for British Teachers Ltd. (Unpublished document)

Hayes, A., 1983: "Planning a project". In Brumfit, pp.15-27.

Kennedy, C. J. , 1988: "Evaluation of the management of change in ELT projects", *Applied Linguistics* 9.4, pp.329-342.

Mackay, R., and A. Mountford (eds), 1977: *English for Specific Purposes: A Case Study Approach*. London: Longman.

Murphy, D. F., 1985: "Evaluation in language teaching: assessment, accountability and awareness". In Alderson, pp.1-17.

Potts, P. J., 1985: "The role of evaluation in a communicative curriculum, and some consequences for materials design". In Alderson, pp.19-44.

Rea, P. M., 1983: "Evaluation of educational projects, with special reference to English language education". In Brumfit, pp.85-98.

Rea, P.M., 1987: "Communicative curriculum validation: a task-based approach". In C. N. Candlin and D. F. Murphy (eds), *Language Learning Tasks*. Lancaster Practical Papers in English Language Education 7. Prentice Hall International and Lancaster University

Roe, P., 1980: "The English Language Centre, Jeddah". In Smyth, pp.172-189.

Smyth, E., (ed.): *Projects in Materials Design*. ELT Documents Special. London: British Council.

Wilson, P., and I. Harrison, 1983: "Materials design in Africa with particular reference to the Francophone Primary School Project, Cameroon". In Brumfit, pp.29-48.

© University of Dublin
Trinity College

Spring 1986

14. D. G. Little and A. J. Grant, *Learning German without a teacher. Report on a self-instructional programme for undergraduate students of Engineering Science at Trinity College, Dublin, 1982-1984* (60pp.)
15. Máire Owens, *Eithne: a study of second language development* (62pp.)
16. D. Wilson and D. Sperber, *Pragmatics: an overview* (36pp.)

Autumn 1986

17. Ailbhe Ní Chasaide& Eugene Davis, *A data-processing system for quantitative analysis in speech production* (28pp.)
18. Seán M. Devitt, *Learning a foreign language through the media* (69pp.)
19. Meriel Bloor & Thomas Bloor, *Languages for specific purposes: practice and theory* (34pp.)

Spring 1988

20. D. G. Little & D. M. Singleton, *Authentic materials and the role of fixed support in language teaching: towards a manual for language learners* (26pp.)

Spring 1989

21. Seán M. Devitt, *Classroom discourse: its nature and its potential for language learning* (72pp.)
22. V. J. Cook, *The relevance of grammar in the applied linguistics of language teaching* (43pp.)

Spring 1990

23. Sera De Vriendt & Pete Van de Craen, *Bilingualism in Belgium: a history and an appraisal* (52pp.)
24. David Singleton, *The cross-linguistic factor in second language learning: a report on some small-scale studies recently conducted at the CLCS* (20pp.)

Autumn 1990

25. Paul Farrell, *Vocabulary in ESP: a lexical analysis of the English of electronics and a study of semi-technical vocabulary* (82pp.)
26. David Singleton, *The TCD Modern Languages Research Project: objectives instruments and preliminary results* (19pp.)
27. Federica Scarpa, *Contrastive analysis and second language learners' errors: an analysis of C-test data elicited from beginners in Italian* (47pp.)

Spring 1991

28. Jennifer Ridley, *Strategic competence in second language performance: a study of four advanced learners* (95pp.)
29. Susan Abbey, *A case for on-going evaluation in English language teaching projects* (43pp.)